

What is claimed is:

1. A method of data mining protein data, the method comprising:
 - accessing data identifying respective outcomes associated with a set of proteins subjected to a set of conditions; and
 - analyzing the data based on the outcomes.
2. The method of claim 1, wherein one of the outcomes comprises identification of protein crystallization of one of the set of proteins in one of the set of conditions.
3. The method of claim 1, wherein one of the outcomes comprises identification of protein solubility of one of the set of proteins in one of the set of conditions.
4. The method of claim 1, wherein at least one of the conditions comprises a solution.
5. The method of claim 1, wherein analyzing comprises determining the efficiency of a set of the conditions in producing a selected outcome in multiple ones of the proteins.
6. The method of claim 5, wherein the multiple ones of the proteins comprises a subset of the set of proteins.
7. The method of claim 6, further comprising selecting the subset of the set of proteins.
8. The method of claim 7, wherein selecting comprises selecting based on the similarity of characteristics of a protein with characteristics of proteins in the set of proteins.

9. The method of claim 5, further comprising determining a prioritized set of conditions.

10. The method of claim 5, further comprising providing a kit on conditions based on the determining.

11. The method of claim 1,
further comprising accessing data identifying
characteristics of the protein; and
wherein analyzing the data comprising analyzing the data
based on the data identifying characteristics of the protein.

12. The method of claim 11, wherein the characteristics
comprise measured characteristics.

13. The method of claim 12, wherein the measured
characteristics comprise at least one of the following: pI,
secondary structure, amino-acid composition, oligometric state,
protein mass, and protein mono-dispersity.

14. The method of claim 11, wherein the characteristics
comprise determined characteristics.

15. The method of claim 14, wherein the determined
characteristics comprise at least one of the following: protein
sequence, amino acid composition, predicted pI, net charge, ratio
of one or more pairs of amino acids, mass, predicted secondary
structure, and predicted tertiary structure.

16. The method of claim 11, wherein the characteristics
comprise an encoding of the 3D structure of the protein.

17. The method of claim 11, wherein the characteristics comprise identification of the concentration of the protein.

18. The method of claim 11, wherein the characteristics comprise identification of a function of the protein.

19. The method of claim 11, wherein the characteristics comprise at least one location of the protein.

20. The method of claim 11, wherein the characteristics comprise additives to the protein.

21. The method of claim 11,
further comprising accessing data identifying
characteristics of different ones of the conditions; and
wherein analyzing the data comprising analyzing the data
based on the data identifying characteristics of the conditions.

22. The method of claim 21, wherein the condition
characteristics comprise pH.

23. A computer program product, disposed on a computer
readable medium, for data mining protein data, the computer
program product including instructions for causing a processor
to:

access data identifying respective outcomes associated with
a set of proteins subjected to a set of conditions; and
analyze the data based on the outcomes.

24. The computer program of claim 23, wherein one of the
outcomes comprises identification of protein crystallization of
one of the set of proteins in one of the set of conditions.

25. The computer program of claim 23, wherein one of the outcomes comprises identification of protein solubility of one of the set of proteins in one of the set of conditions.

26. The computer program of claim 23, wherein the instructions for causing the processor to analyze comprise instructions for causing the processor to determine the efficiency of a set of the conditions in producing a selected outcome in multiple ones of the proteins.

27. The computer program of claim 26, wherein the multiple ones of the proteins comprises a subset of the set of proteins.

28. The computer program of claim 27, further comprising instructions for causing the processor to select the subset of the set of proteins.

29. The computer program of claim 28, wherein the instructions for causing the processor to select comprise instructions for causing the processor to select based on the similarity of characteristics of a protein with characteristics of proteins in the set of proteins.

30. The computer program of claim 23,
further comprising instructions for causing the processor to access data identifying characteristics of the protein; and
wherein the instructions for causing the processor to analyze the data comprise instructions for causing the processor to analyze the data based on the data identifying characteristics of the protein.

31. The computer program of claim 23,

further comprising instructions for causing the processor to access data identifying characteristics of different ones of the conditions; and

wherein the instructions for causing the processor to analyze the data comprise instructions for causing the processor to analyze the data based on the data identifying characteristics of the conditions.

PCT/US2014/042205